

DOCUMENT RESUME

ED 167 574

TM 008 020

AUTHOR Fletcher, Jerry L.
 TITLE Improving the Availability of Data in Education: The Federal Role.
 PUB DATE 29 Mar 78
 NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)
 EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
 DESCRIPTORS *Data Bases; Data Collection; *Educational Assessment; Educational Research; *Evaluation Methods; *Evaluation Needs; Evaluators; Federal Government; *Futures (of Society); Government Role; *Program Evaluation; Research Utilization

ABSTRACT

Questions are raised about the future of research and development evaluation. Without a set of overarching policy questions or theoretical questions, evaluation data are simply not cumulative and cannot be organized into any data base. Education has never been a data-based enterprise. The meaning of evaluation should be broadened to include the generation of systematically collected information relevant to policy decisions about programs, and the development of models to improve understanding of problems and interventions. Evaluation should adapt to the life cycle of the program. A series of small, single-objective, studies may be more useful than a single large and complex study. The plans for evaluation should be negotiated with all the interested parties and the results should be addressed to the interests of all these groups and effectively synthesized. The federal government is likely to support large data banks and efforts to synthesize evaluation data, and to reduce support for evaluation of separate projects.
 (Author/CTM)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Gerry S. Fletcher

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM "

U. S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY.

ABSTRACT

Improving the Availability of Data in Education: The Federal Role

This paper suggests that a new approach to evaluation is about to come into being, based on the lessons of the past ten years. The definition of evaluation will be broadened to include many types of systematic data collection, particularly efforts to understand why effects are or are not created; different types of studies will be done at different stages in the life of a program; the kinds of data to be collected will represent a negotiated agreement among the many interested parties; and the data will be presented in interpretative syntheses of results. The government can be expected to move toward assembling evaluation data into large data banks, and making it widely available for interpretive syntheses. Within this framework, R&D evaluation is likely to become anachronistic, as it is not yet amenable to assembling in large data-banks.

ED167574

TM008 020

IMPROVING THE AVAILABILITY OF DATA IN EDUCATION: THE FEDERAL ROLE

The views in this paper are those of the author and no endorsement by HEW or the Federal government should be inferred or implied.

by: Jerry L. Fletcher
Senior Policy Analyst
Office of Assistant Secretary
for Education/Policy Development
200 Independence Avenue, SW
Washington, D.C. 20201
(202) 245-8266

Prepared for delivery at the Annual Conference: American Educational
Research Association, Toronto, Canada March 29, 1978 Session 18.18

PRINTED IN U.S.A.

For the past two years, I have written a regular column in the newsletter of special interest group organizing this session: R & D Evaluators; and in that time have simultaneously been involved in working with evaluation at the Federal level, watching and being involved in efforts to pull together evaluation information into some useful form to guide policy decisions. In the course of the last two years I have gone through several different stages of depression about it all, trying to get ahold of why R & D evaluation seemed so fundamentally irrelevant to anything I was doing. Several of the columns of the newsletter have dealt with this in some dimension, though there has rarely been any reaction from people, so I'm not sure whether it had any effect. In the past few months, as I thought about this paper, some of the ideas and experiences have jelled, and this paper is an attempt to formulate them.

Education has never been a data-based enterprise. My initial job after I got my doctorate was as the director of research in an experimental high school. While I was able to a certain degree to gather information that should have been part of discussions about what directions and changes the school should take, it was immediately evident that there was no place to insert this infor-

mation into the ongoing life of the school. I literally had to create a set of mechanisms for stopping the school's headlong day-to-day rush for awhile to get people to consider the information I had collected.

After I left the school system, I became head of research and evaluation for a very large development effort at one of the Regional Laboratories. It was clear to me then, and it is clear to me now, that educational development also has never been a data-based enterprise. Since most of the people doing the work are out of the tradition of education, there is little in the way of a commitment to creating the ways of operating that would enable a careful consideration of systematically collected information about the relative effectiveness of what is being developed. As with teachers in classrooms, ones that our tradition say are the best, one has an idea, tries it, gets a sense of how it works, revises it, etc. until the product is developed. Systematically collected information is often seen as irrelevant.

It may have been that I was bad at it, but I doubt that that was the whole story. I personally come out of the tradition of the hard sciences, and my degree was in the history of science. I had a long schooled sense of the way in which genuine inquiry takes place, and it doesn't follow the kinds of logical paradigms that the evaluators who were more naive about the process used to suggest to me. I was supposed to demand that the developers specify their objectives in advance, explain how they intended to create these outcomes, write it all down, allow me to set up experimental design conditions for examining whether they were attaining their objectives, and furthermore, listen to me if I reported that they weren't. The problems

with this are at least two-fold: genuine development can never take place this way; and evaluators who believe that information will in fact determine decisions (i.e., the evidence that I've been listened to is that they make the decisions suggested by the data) are simply ignorant of the epistemological and philosophical basis of what they do. When pressed, others would always mouth the statement that, of course, data didn't make decisions, but the underlying way we justified our reason for being was by pointing to instances where it did.

In the past two years at the Federal level, there has developed an increasing sophistication about the relationship between evaluation and decision-making, spurred on certainly by the quality of work that John Evans and his staff have done in OE. His work has consistently, in my view, represented what we in the evaluation profession have said was what he ought to do, and as he has had the largest budget and the most visibility, he has taken more than his share of lumps for the weaknesses of our models of evaluation.

Out of this, though, has grown a richer, more complex sense of the relationship between evaluation and the process of improving programs that has implications for all of what we, as evaluators, do. A paper commissioned through the office where I work has just been written, which pulls together all of this thinking as clearly as I have seen anywhere. It is written by Sue Berryman and Tom Glennan, both of the RAND Corporation, and will be available soon. I believe it to be the framework for where evaluation needs to go next.

The paper is written with respect to the evaluation of large-scale Federal programs, but the analysis has great power for reorganizing evaluation at every level. I will review some of the highlights of their paper, but explain them with reference to R&D evaluation. I will then conclude by suggesting some of the implications this new framework has.

First, they recommend broadening the definition of evaluation so that it is not just "placing value" on something, or knowing the effects of a program, but "generating information relevant to policy decisions about programs." In other words, it is a more useful way of defining evaluation to have it refer to "systematically collected information" than to any particular kind of information such as "effects" or "outcome" information.

Second, they argue that evaluation studies are much too wedded to experimental designs as ways of drawing inferences from effects data. They argue that we also need to include methods of "understanding" the problem, the intervention, and the major events which can affect the interpretation of the results. "The effects of a program cannot be adequately interpreted on the basis of effects data alone." Evaluators need to get involved with models, with why something does or does not work. Too often we have null, or very small positive effects, and no idea what this means we should do.

Third, they argue that any program has a life cycle, and that the nature of the studies undertaken with regard to a program should vary through the program's life cycle. While they are speaking particularly of

a new Federal program, what they say is true for development programs. The kinds of questions that ought to be asked in the early stages of a program, the way they ought to be asked, the degree to which certain methodologies are appropriate, all need to be reexamined. For example, they suggest that the first set of studies for any new program ought to be concerned with whether its objectives are realistic. Only later in the life cycle, after it is clear that there is a treatment and that it has reasonably well stabilized, does it make sense to ask about the program's effects. At best in the early phases there ought to be studies of the feasibility of obtaining effects data later.

Fourth, they argue for a series of small, single objective studies, undertaken in such a way that the subsequent studies can follow up on leads suggested by the earlier studies. Too often there is a large study, with many objectives, and it is very difficult to adjust it to pursue leads that seem to provide the most relevant information.

They also make a major contribution to the way we think about the audiences of evaluation information. Again, while they are talking about Federal programs, the points they make are relevant to R&D evaluation, and I will speak of them in that framework. They argue, correctly, that information does not make decisions. There is a "political" process which determines which of the many decision paths will be taken, regardless of the data. In almost any setting, and our R&D Centers or Labs are no exception, there are often many parties who have some "political" interest in what decision is made: the relevant parties for a normal Federal contract might be the Federal program officer, the administration of the organization

holding the contract, the particular project director at the development center, and possibly as a separate interest group the particular workers working on some phase of the overall project.

In the Bergman-Glennan view, a plan for evaluation studies should be negotiated between the evaluation staff and these interest groups. It should be a multi-year plan, revised annually. Furthermore, the negotiation should be of the form: what kinds of information can we all agree would be valuable for us to have, in making decisions relevant to this program. Once this is established, with whatever horse-trading, log rolling, or whatever else takes place, the methodology should be left to the evaluators, and they should be as independent of bias as it is possible for them to be, so that all parties to the decision-making will at least trust the quality of the data, even if in the last analysis they choose to disregard it.

Finally, they suggest that reporting the results of evaluation studies is not the form that makes them usable by decision-makers. They suggest that results should be reported in the form of syntheses of all of the findings relevant to specific policy issues. The format should be more nearly that of a good staff report, rather than that of a conventional research paper. Put in simple terms, data from a whole variety of sources ought to be accumulated in some organized way so that syntheses can be generated relevant to any major issues that the decision-makers face. And furthermore, evaluators should make these organized data-bases available to others so that pluralistic sources of synthesis can be engaged.

These points: a broadened definition of evaluation; the use of

methods of understanding why certain effects are or are not created; the adjustment of evaluation methods to the life cycle of a program; the sequential series of small studies; the negotiated evaluation plan; and the preparation of interpretative syntheses from all of the available data represent in combination a good picture of the way evaluations will increasingly be conducted in the future.

Moving in these directions will require some major changes at the Federal level. Education ranks near the bottom in terms of the quality and accessibility of data, compared, lets say, to the Bureau of Labor Statistics, or the data on health, housing, GNP, business, or cost-of-living. Within the educational establishment as a whole, enormous amounts of data are collected, used once or twice, and discarded or forgotten. Within various operating institutions, from individual schools to universities, large amounts of data which might be used to answer critical questions are routinely collected, but no one bothers to assemble the data, organize them into accessible data bases, make the data into useful indices, or otherwise shape it for wider use.

In the next few years I believe we can anticipate that the National Center for Educational Statistics will move dramatically in these directions, providing for education regular reports on various statistical indicators, providing to researchers regularly updated data bases. What presently is an annoyance of forms clearance will, I believe, develop into a genuine attempt to make all of the data collected anywhere in the education enterprise potentially useable over long periods of time by being incorporated into large data bases. And, I suspect the government will increasingly support multiple syntheses

of the relevant information around policy questions of great importance.

Within this context it is hard to know whether R&D evaluation as we presently know it will be little more than an anachronism. There is much that can be learned from the analysis of the Berryman-Glennan piece that would immensely improve the quality and utility of the R&D evaluations presently conducted, but the fact remains that most of these are specific studies of relatively idiosyncratic development efforts, and we have not developed within our professional group any sense of a set of larger questions, across many development efforts, to which our individual studies could be related. Without such a set of overarching questions--either related to the policy questions about whether more development should be funded; or to some theoretical base, such as a theory of development work, the data is simply not cumulative. It cannot be organized into any data base. It cannot be interrogated to find information relevant to any known policy questions, except those facing the particular program and the particular developers.

While it still may be in the interest of the government to provide some small amount of money for such particularistic evaluations, as an overall management tool, I suspect those engaged in R&D evaluation will increasingly be the forgotten people of evaluation work, laboring on tiny budgets attempting to satisfy themselves that what they do contributes something to the quality of a string of isolated projects as they move from one to another through their careers. I don't like it much as a vision of what practicing evaluation means, though I think it is upon us more than many of us realize. From time to time the best of our profession happen through Washington, and those who stop to see me always seem harried by the demands

of running from one small budget evaluation to another. Few of them have time to write. Fewer still time to think. And one even said to me recently that it was worse if he did think about it, because he no longer had any faith in what he was doing, but he still had to go ahead and do it.

If this special interest group is to provide any leadership to the profession, I suggest it lies in finding ways to do the thousands of individual evaluation studies so that the data can be organized and interrogated to answer larger questions. It remains to be seen if we will move to establish some principles for our work that will accomplish this goal.